

Identification and Removal of Noise Modes in *Kepler* Photometry

Erik A. Petigura

Astronomy Department, University of California, Berkeley, CA 94720

epetigura@berkeley.edu

Geoffrey W. Marcy

Astronomy Department, University of California, Berkeley, CA 94720

ABSTRACT

We present the Transiting Exoearth Robust Reduction Algorithm (TERRA) — a novel framework for identifying and removing instrumental noise in *Kepler* photometry. We identify instrumental noise modes by finding common trends in a large ensemble of light curves drawn from the entire *Kepler* field of view. Strategically, these noise modes can be optimized to reveal transits having a specified range of timescales. For *Kepler* target stars of low photometric noise, TERRA produces ensemble-calibrated photometry having 33 ppm RMS scatter in 12-hour bins, rendering individual transits of earth-size planets around sun-like stars detectable as $\sim 3\sigma$ signals.

Subject headings: techniques: photometric — planetary systems

1. Introduction

The *Kepler* Mission is ushering in a new era of exoplanet science. Landmark discoveries include *Kepler*-10b, a rocky planet (Batalha et al. 2011); the *Kepler*-11 system of six transiting planets (Lissauer et al. 2011); earth-sized *Kepler*-20e and 20f (Fressin et al. 2012); KOI-961b, c, and d — all smaller than earth (Muirhead et al. 2012); and *Kepler*-16b a circumbinary planet (Doyle et al. 2011). While *Kepler* has revealed exciting individual systems, the mission’s legacy will be the first statistical sample of planets extending down to earth size and out to 1 AU. *Kepler* is the first instrument capable of answering “How common are earths?” — A question that dates to antiquity.

Planet candidates are detected by a sophisticated pipeline developed by the *Kepler* team Science Operations Center. In brief, systematic effects in the photometry are suppressed by the Pre-search Data Conditioning (PDC) module, the output of which is fed into the Transiting Planet Search (TPS) module. For further information, see Jenkins et al. (2010).

The *Kepler* mission was designed to study astrophysical phenomena with a wide range of timescales, which include 1-hour transits of hot Jupiters, 10-hour transits of planets at 1 AU, and

weeklong spot modulation patterns. The PDC module is charged with removing instrumental noise while preserving signals with a vast range of timescales. We review sources of instrumental errors in § 2, highlighting the effects that are most relevant to transit detection.

The *Kepler* team has released candidate planets based on the first 4 and 16 months of data (Borucki et al. 2011; Batalha et al. 2012). Many of the candidates have additional followup observations from the ground and space aimed at ruling out false positive scenarios. In addition, statistical arguments suggest that 90-95% of all candidates and that $\sim 98\%$ of candidates in multi-candidate systems are bonafide planets (Morton & Johnson 2011; Lissauer et al. 2012).

While *Kepler*’s false positive rate is low, its completeness is largely uncharacterized. If the completeness decreases substantially with smaller planet size or longer orbital periods, the interpretations regarding occurrence drawn from the Borucki et al. (2011) and Batalha et al. (2012) catalogs will be incorrect. Hunting for the smallest planets, including earth-sized planets in the habitable zone, will require exquisite suppression of systematic effects. Without optimal detrending, systematic noise will prevent the detection of the smallest planets, possibly the habitable-zone earth-sized planets, which is the main goal of the *Kepler* mission. Therefore, it is essential for independent groups to develop pipelines that compliment both PDC and TPS. An early example of an outside group successfully identifying new planet candidates is the Planet Hunters project (Fischer et al. 2011; Lintott et al. 2012), which uses citizen scientists to visually inspect light curves. In addition, existing pipelines from the HAT ground-based search (Huang et al. 2012) and the *CoRoT* space mission (Ofr & Dreizler 2012) have been brought to bear on the *Kepler* dataset yielding ~ 100 new planet candidates.

We present the Transiting Exoearth Robust Reduction Algorithm (TERRA) — a framework for identifying and removing systematic noise. We identify systematic noise terms by searching for photometric trends common to a large ensemble of stars. Our implementation is tuned toward finding trends with transit-length timescales.

2. Instrumental Noise in *Kepler* Photometry

The *Kepler* spacecraft makes photometric observations of $\sim 156,000$ targets. Long cadence photometry is computed by summing all the photoelectrons within a predefined target aperture during a 29.4 minute integration. The *Kepler* team makes this “Simple Aperture Photometry” available to the scientific community (Fraquelli & Thompson 2012). Simple aperture photometry contains many sources of noise other than Poisson shot noise. We illustrate several noise sources in Figure 1, where we show the normalized photometry (δF) of KIC-8144222 ($Kp = 12.4$). $\delta F = (F - \bar{F})/\bar{F}$ where F is the simple aperture photometry.

The dominant systematic effect on multi-quarter timescales is “differential velocity aberration” (Van Cleve & Caldwell 2009). As *Kepler* orbits the sun, its velocity relative to the *Kepler* field changes. When the spacecraft approaches the *Kepler* field, stars on the extremities of the field

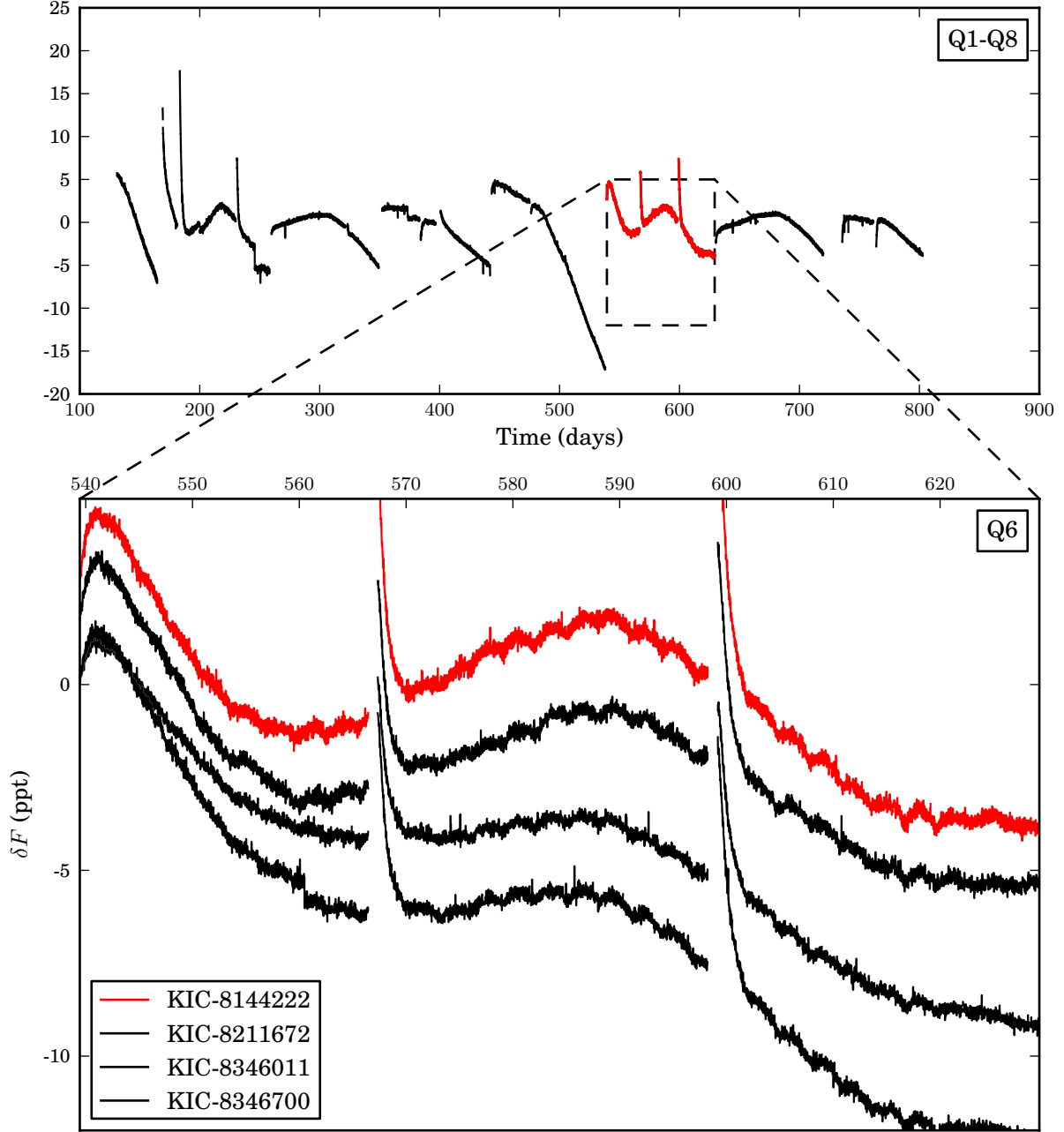


Fig. 1.— Top: Normalized flux from KIC-8144222 ($Kp = 12.4$, $CDPP12 = 35.4$ ppm) from Quarter 1 through 8 (Q1-Q8). Bottom: Detail of Q6 photometry showing KIC-8144222 along with three stars of similar brightness, noise level, and location on the FOV ($12.0 < Kp < 13.0$, $CDPP12 < 40$ ppm, $\text{mod.out} = 16.1$). Much of the variability is common to the 4 stars and therefore instrumental in origin. The two spikes are due to thermal settling events, and the three-day ripples are due to onboard momentum management.

move toward the center. Stellar PSFs move over *Kepler* apertures by ~ 1 arcsecond resulting in a ~ 1 % effect over 1-year timescales.

We show a detailed view of KIC-8144222 photometry from Quarter 6 (Q6) in Figure 1. The decaying exponential shapes are caused by thermal settling after data downlinks. Each month, *Kepler* rotates to orient its antenna toward earth. Since *Kepler* is not a uniformly colored sphere, changing the spacecraft orientation with respect to the sun changes its overall temperature. After data downlink, *Kepler* takes several days to return to its equilibrium temperature (Jeffrey Smith, private communication, 2012). KIC-8144222 photometry also shows a $\sim 0.1\%$ effect with a 3-day period due to thermal coupling of telescope optics to the reaction wheels. We explore this 3-day cycle in depth in § 3.3.

Since all of the previously mentioned noise sources are coherent on timescales longer than one cadence (29.4 minutes), the RMS of binned photometry does not decrease as $1/\sqrt{N}$, where N is the number of measurements per bin. In order to describe the noise on different timescales, the *Kepler* team computes quantities called CDPP3, CDPP6, and CDPP12 which are measures of the photometric scatter in 3, 6, and 12-hour bins. KIC-8144222 has CDPP12 35.4 ppm and is a low-noise star (bottom 10 percentile). For a more complete description of noise in *Kepler* data see Christiansen et al. (2011).

As a comparison, we selected stars which were similar to KIC-8144222 in position on the Field of View (FOV), noise level, and brightness ($\text{mod.out} = 16.1$, $\text{CDPP12} < 40$ ppm, $12.0 < Kp < 13.0$). From this 13-star sample, we randomly selected 3 stars and show their light curves in Figure 1. The photometry from the comparison stars is strikingly similar to the KIC-8144222 photometry. Since much of the variability is correlated, it must be due to the state of the *Kepler* spacecraft. Common trends among stars can be identified and removed. The *Kepler* team calls this “cotrending,” a term we adopt.

Correlated noise with timescales between 1 and 10 hours can mimic planetary transits and requires careful treatment. To illustrate the transit-scale correlations among a large sample of stars, we show a correlation matrix constructed from 200 Q6 light curves in Figure 2. The *Kepler* photometer is an array of 42 CCDs arranged in 21 modules (Fraquelli & Thompson 2012). We organized the rows and columns of the correlation matrix by module. We constructed the correlation matrix using the following steps:

1. We randomly selected 10 light curves from each of the 20 total modules¹ from stars with the following properties: $12.5 < Kp < 13.5$ and $\text{CDPP12} < 40$ ppm.
2. To highlight transit-scale correlations, we subtracted a best fit spline from the photometry. The knots of the spline are fixed at 10-day intervals so that we remove trends $\gtrsim 10$ days.

¹Module 3 failed during Q4 (Christiansen et al. 2011).

3. We normalized each light curve so that its median absolute deviation (MAD) is unity.
4. We evaluated the pairwise correlation (Pearson-R) between all 200 stars.

The correlation matrix shows that stars in some modules (e.g. module 2) correlate strongly with other stars in the same module. However, other modules (e.g. module 12) shows little inter-module correlation. Finally, the large off-diagonal correlations show that stars in some modules correlate strongly with stars in different modules.

3. Identification of Photometric Modes

We have shown that there is significant high-frequency ($\lesssim 10$ days) systematic noise in *Kepler* photometry. In order to recover the smallest planets, this noise must be carefully characterized and removed. We isolate systematic noise by finding common trends in a large ensemble of stars. This is an extension of differential photometry, widely used by ground-based transit surveys to calibrate out the time-variable effects of the earth’s atmosphere. We find these trends using Principle Component Analysis (PCA). This is similar to the Sys-Rem, TFA, and PDC algorithms (Tamuz et al. 2005; Kovács et al. 2005; Twicken et al. 2010), but our implementation is different. We briefly review PCA in the context of cotrending a large ensemble of light curves.

3.1. PCA on Ensemble Photometry

Consider an ensemble of N light curves each with M photometric measurements. We can think of the ensemble as a collection of N vectors in an M -dimensional space. Each light curve δF can be written as a linear combination of M basis vectors that span the space,

$$\begin{aligned} \delta F_1 &= a_{1,1}V_1 + \dots + a_{1,M}V_M \\ &\vdots \\ \delta F_N &= a_{N,1}V_1 + \dots + a_{N,M}V_M \end{aligned} \tag{1}$$

where each of the V_j basis vectors is the same length as the original photometric time series. Equation 1 can be written more compactly as

$$\mathbf{D} = \mathbf{A}\mathbf{V}$$

where

$$\mathbf{D} = \begin{pmatrix} \delta F_1 \\ \vdots \\ \delta F_N \end{pmatrix}, \mathbf{A} = \begin{pmatrix} a_{1,1} & \dots & a_{1,M} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \dots & a_{N,M} \end{pmatrix}, \mathbf{V} = \begin{pmatrix} V_1 \\ \vdots \\ V_M \end{pmatrix}$$

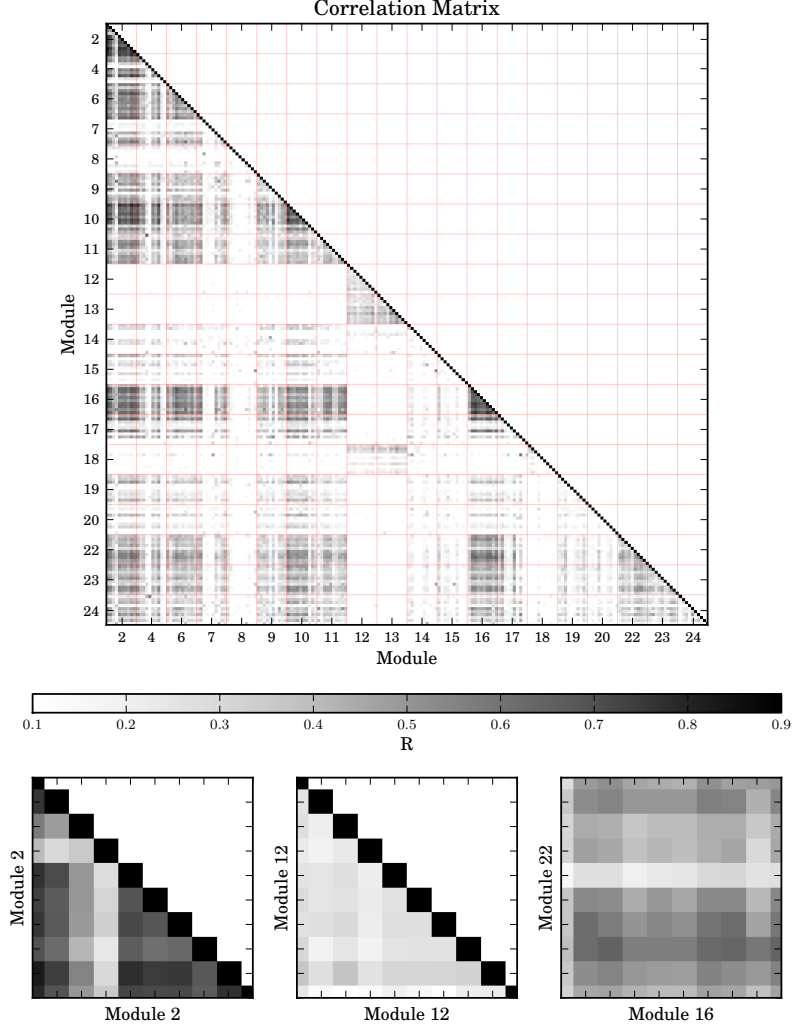


Fig. 2.— Top: Correlation matrix constructed from 200 Q6 light curves. The correlation (R-value) between two stars is represented by the gray scale, which ranges from 0.1 to 0.9. The diagonal elements have $R = 1$. The stars are ordered according to module and the red lines delineate one module from another. We enlarge several 10x10 regions in the lower panels. Stars in some modules (such as module 2) are highly correlated, while other modules (such as module 12) show little correlation. The module 22 - module 16 correlation matrix is an example of significant intra-module correlation.

Singular Value Decomposition (SVD) simultaneously solves for the basis vectors \mathbf{V} and the coefficient matrix \mathbf{A} because it decomposes any matrix \mathbf{D} into

$$\mathbf{D} = \mathbf{U}\mathbf{S}\mathbf{V}^T.$$

\mathbf{V} is an $M \times M$ matrix where the columns are the eigenvectors of $\mathbf{D}^T\mathbf{D}$ or “principle components,” and the diagonal elements of \mathbf{S} are the corresponding eigenvalues. The eigenvalues $\{s_{1,1}, \dots, s_{M,M}\}$ describe the extent to which each of the principle components capture variability in the ensemble and are ordered from high to low. The columns of \mathbf{U} are the eigenvectors of $\mathbf{D}\mathbf{D}^T$. Both \mathbf{U} and \mathbf{V} are unitary matrices, i.e. $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ and $\mathbf{V}\mathbf{V}^T = \mathbf{I}$.

As we saw in § 2, stars show common photometric trends due to changes in the state of the *Kepler* spacecraft. The most significant principle components will correspond to these common trends. If we identify the first N_{Mode} principle components as instrumental noise modes, we can remove them via

$$\delta F_{i,cal} = \delta F_i - \sum_{j=1}^{N_{Mode}} a_{i,j} V_j \quad (2)$$

where δF_{cal} is an ensemble-calibrated light curve. However since the collection of $\{V_i, \dots, V_M\}$ spans the space, the higher principle components describe astrophysical variability, shot noise, and exoplanet transits. We must be careful not to remove too many components because we would be removing the signals of interest.

3.2. PCA implementation

We construct a large reference ensemble of light curves $\{\delta F_1, \dots, \delta F_N\}$ of 1000 stars ($12.5 < Kp < 13.5$, $CDPP12 < 40$ ppm) drawn randomly from the entire FOV. Before performing SVD, we remove thermal settling events and trends $\gtrsim 10$ days as described in § 2. Since SVD finds the eigenvectors of $\mathbf{D}^T\mathbf{D}$ it is susceptible to outliers as is any least squares estimator. We perform a robust SVD that relies on iterative outlier rejection following these steps:

1. Find principle components and weights for light curve ensemble.
2. The i^{th} light curve is considered an outlier if any of the mode weights $(a_{i,1}, \dots, a_{i,4})$ differ significantly from the typical mode weight in the ensemble. We consider $a_{i,j}$ to be significantly different from the ensemble if

$$\frac{|a_{i,j} - \text{med}(a_j)|}{\text{MAD}(a_j)} > 10$$

where $\text{med}(a_j)$ and $\text{MAD}(a_j)$ are the median value and the median absolute deviation of all the a_j mode weights.

3. Remove outlier light curves from the ensemble.

4. Repeat until no outliers remain.

For our 1000-star sample we identified and removed 51 stars from our ensemble. These stars tended to have high amplitude intrinsic astrophysical variability, i.e. due to spots and flares. We plot the four most significant TERRA principle components in Figure 3 and offer some physical interpretations of the mechanisms behind these modes in the following section.

3.3. Interpretation of Photometric Modes

In this section, we associate the variability captured in the principle components to changes in the state of the *Kepler* spacecraft that couple to photometry. The three-day cycle isolated in our first principle component is due to a well-known, three-day momentum management cycle on the spacecraft (Christiansen et al. 2011). To keep a fixed position angle, *Kepler* must counteract external torques by spinning up reaction wheels. These reaction wheels have frictional losses which leak a small amount of heat into the spacecraft, which changes the PSF width and shape of the stars.

We can gain a more detailed understanding of this effect, by examining how the mode weights for each reference star corresponding to V_1 , i.e. $\{a_{1,1}, \dots, a_{N,1}\}$, vary across the FOV. We display the RA and Dec positions of our 1000-star sample in Figure 4 and color-code the points with the value of a_i . The a_1 and a_2 mode weights show remarkable spatial correlation across the FOV. That a_1 is positive in the center of the FOV and negative at the edges of the FOV means the systematic photometric errors in these two regions respond to the momentum cycle in an anticorrelated sense. The telescope is focused such that the PSF is sharpest at intermediate distances from the center of the FOV. Since stars in the center and on the extreme edges have the blurriest PSFs (Van Cleve & Caldwell 2009), they respond most strongly to the momentum cycle.

The mechanism behind the variability seen in V_2 is less clear. V_2 includes a high frequency component with a period of 1.68 hours. The *Kepler* team has also noticed this periodicity in the pixel scale (Douglas Caldwell, private communication, 2012). A possible explanation is thermal coupling of the telescope optics to a heater that turns off and on with a ~ 20 minute period. The 1.68 hour variability would be an alias of this higher frequency with the observing cadence of 29.4 minutes. The gradient in a_2 across the FOV suggests the heater is coupled to the telescope optics in a tip/tilt rather than piston sense.

The higher-order components a_3 and a_4 do not show significant spatial correlation, which suggests that V_3 and V_4 are not due to changes in the local PSF. Since V_3 and V_4 have a ~ 10 -day timescale, they could be the high frequency component of the differential velocity aberration trend that was not removed by our 10-day spline.

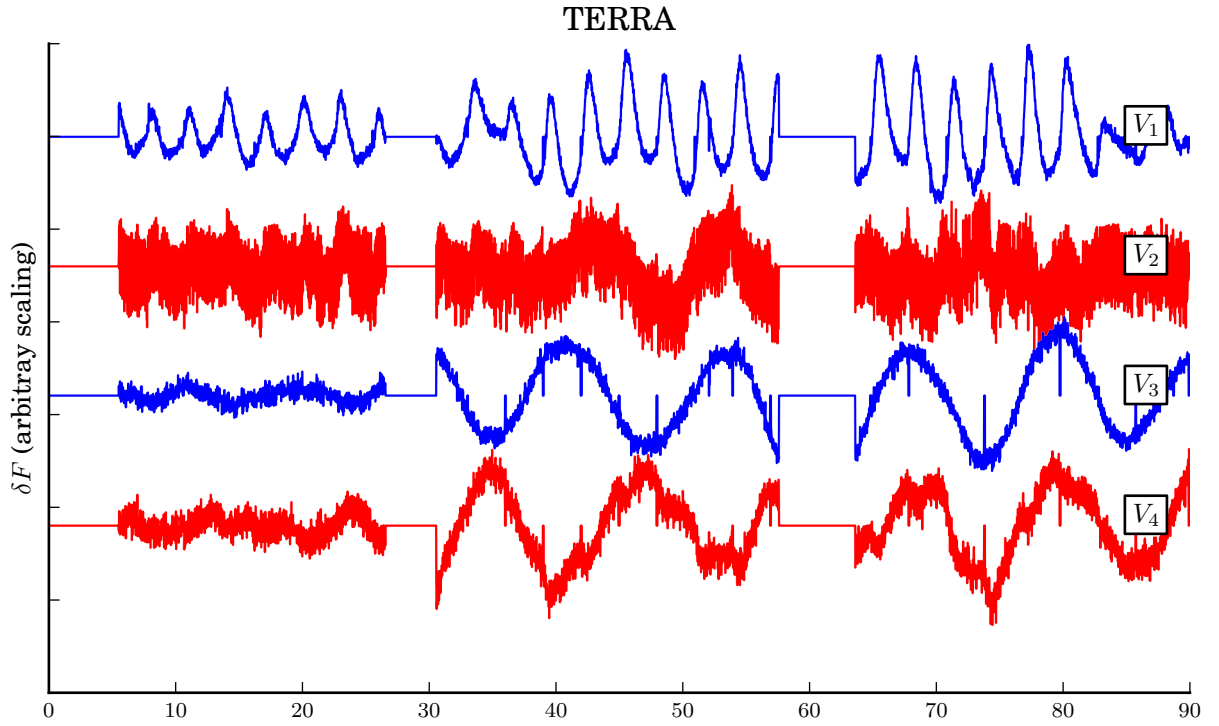


Fig. 3.— Top: The first four TERRA principle components in our 1000-light curve ensemble plotted in order of significance. V_1 has a 3-day periodicity and is due to changes in the thermal state of the spacecraft caused by a 3-day momentum management cycle. V_2 has a high frequency component ($P = 1.68$ hours) that could be due to a 20 minute thermal cycle from an onboard heater aliased with the 29.4 minute observing cadence alias of a 20-minute thermal cycle driven by an onboard heater.

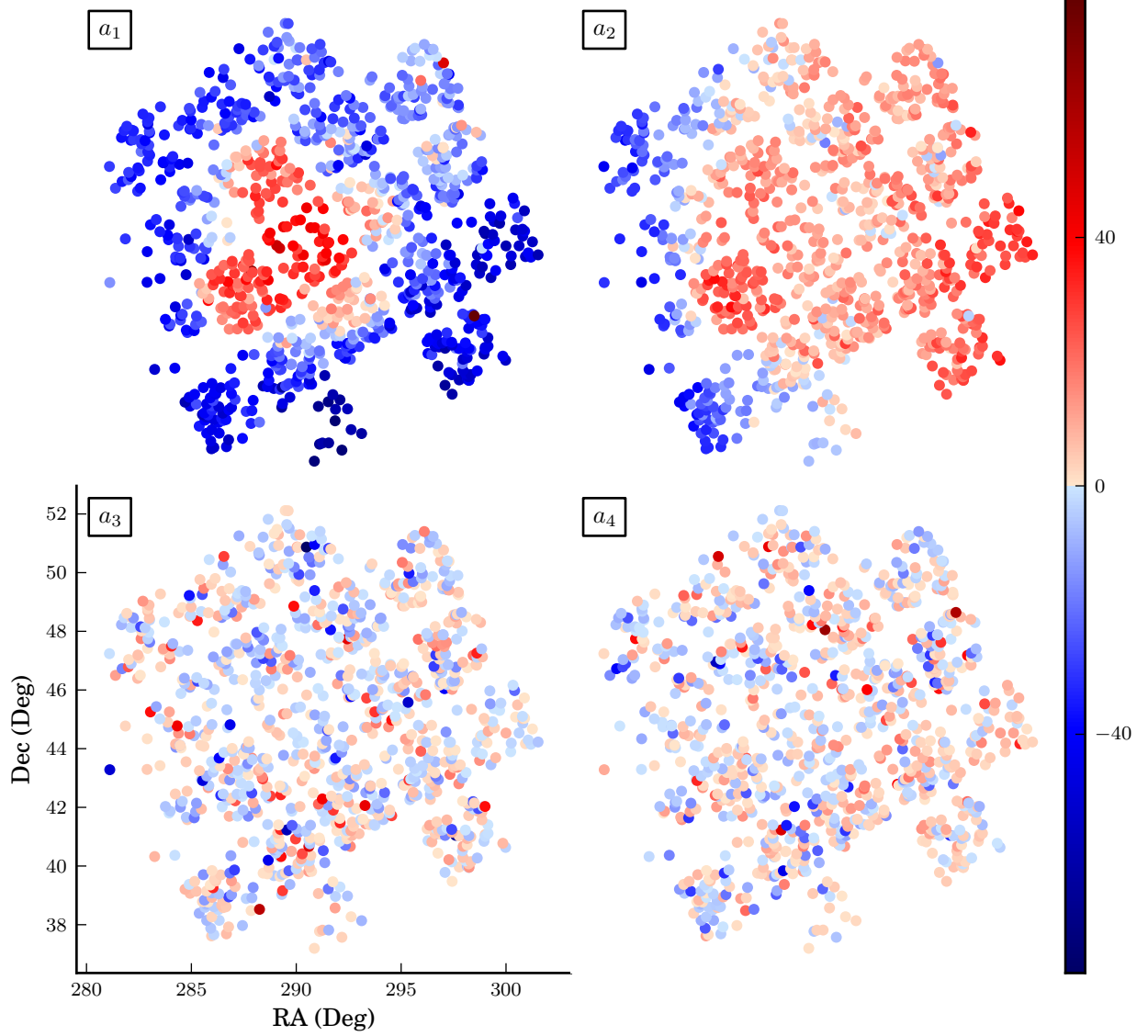


Fig. 4.— The RA and Dec positions of our 1000-star ensemble. The points are color-coded by a_i , the weights for mode V_i . Negative values are shown in blue and positive values are shown in red. The fact that the sign and magnitude of a_1 depends on distance from the center of the FOV supports the idea that the variability captured by V_1 is due to PSF breathing of the telescope which is driven by the three-day momentum management cycle. The gradient in a_2 could be due to the thermal coupling of an onboard heater to the optics in a tip/tilt sense. Mode weights a_3 and a_4 show no spatial correlation and do not seem to depend on changes in the PSF width.

4. Calibrated Photometry

4.1. Removal of Modes

After determining which of the N_{Mode} principle components correspond to noise modes, we can remove them according to Equation 2. In Figure 5, we show fits to KIC-8144222 Q6 photometry using different combinations of TERRA principle components. We achieve uniform residuals using only 2 of our modes as we show quantitatively below. The simplicity of our model buys some insurance against overfitting.

4.2. Performance

For each of the residuals in Figure 5, we computed the mean depth $\overline{\delta F}(t_i)$ of a putative 12-hour transit centered at t_i for every cadence in Q6. The distribution of $\overline{\delta F}$ due to noise determines the minimum transit depth that can be detected by a transit search algorithm. $\overline{\delta F}$ is computed by

$$\overline{\delta F}(t_i) = [\delta F * g](t_i)$$

where ‘*’ denotes convolution and g is the following kernel

$$g(t_i) = \frac{1}{N_T} \left[\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{\text{length} = N_T}, \underbrace{-1, \dots, -1}_{\text{length} = N_T}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{\text{length} = N_T} \right].$$

where N_T is 24. For each of the cotrending schemes, we computed the following statistics describing the distribution of $\overline{\delta F}$: standard deviation (σ), 90 percentile (90 %), and 99 percentile (99 %). The standard deviation is roughly equivalent to CDPP12. Since transit search algorithms key off on peaks in $\overline{\delta F}$, the percentile statistics are more appropriate figures of merit. We list these statistics for KIC-8144222 in Table 1. Ensemble-calibrated photometry produced tighter distributions in $\overline{\delta F}$ than the spline baseline.

4.3. Comparison to PDC

In this section, we offer some simple comparisons between TERRA and the PDC implementation of Twicken et al. (2010). This paper represents our efforts to improve upon that algorithm. The *Kepler* PDC pipeline has evolved beyond that presented in Twicken et al. (2010) culminating with PDC-MAP (Stumpe et al. 2012; Smith et al. 2012). We feel that the Twicken et al. (2010) algorithm is an important touchstone for comparison given that the most recent release of planets (Batalha et al. 2012) was based on photometry that was largely processed with the Twicken et al. (2010) algorithm.

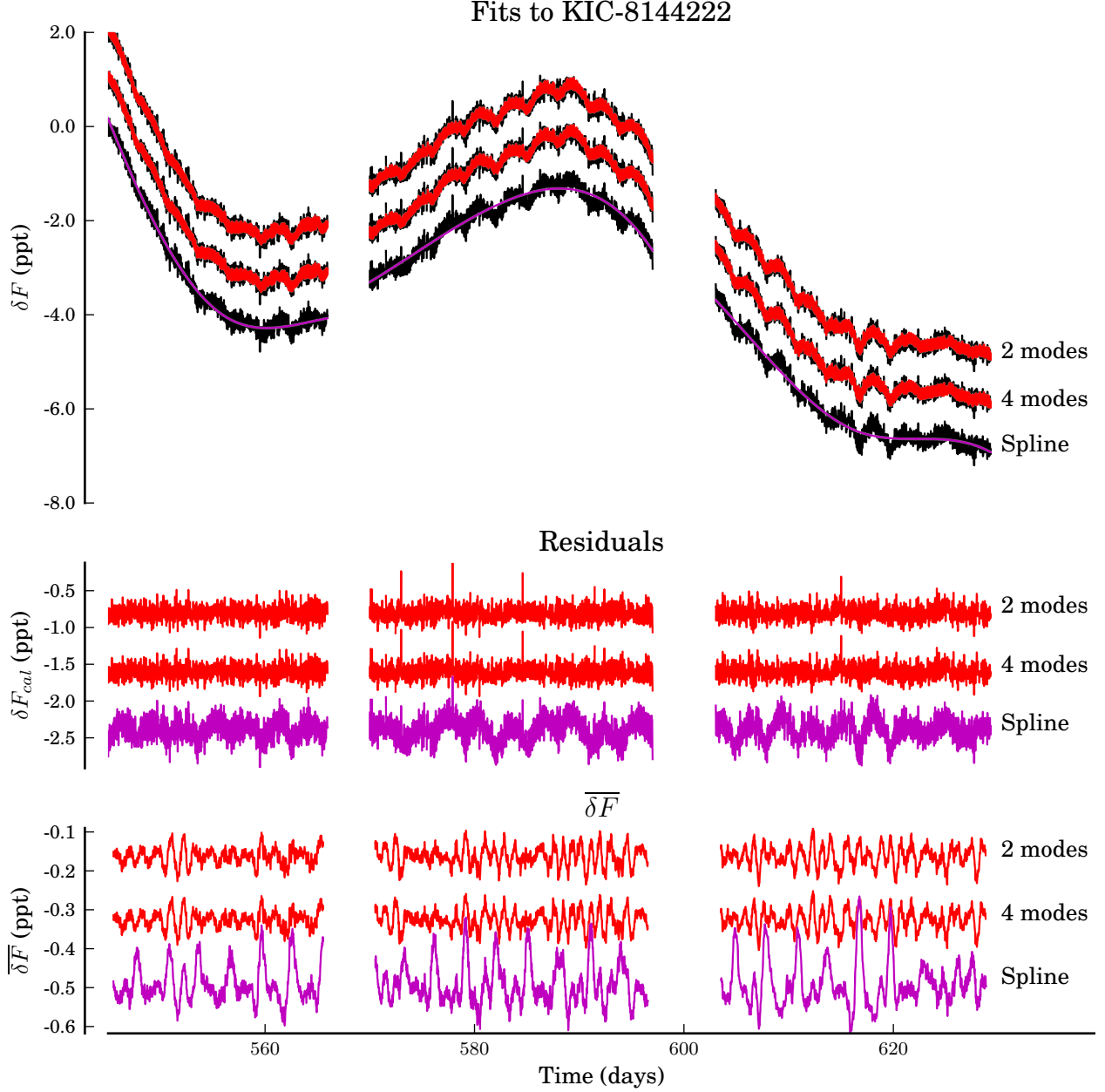


Fig. 5.— Least squares fits using TERRA principle components to KIC-8144222 Q6 photometry. The bottom panel shows 12-hour $\overline{\delta F}$ where smaller scatter implies greater sensitivity to small transits. We show the spline fit (magenta) as a baseline since it incorporates no ensemble-based cotrending information. The $\overline{\delta F}$ using spline detrending shows large spikes at the momentum cycle cusps, which are suppressed in the TERRA cotrending. Using our robust modes, we are able to produce a clean, calibrated light curve using only two modes. Decreased model complexity helps guard against overfitting.

Table 1. Comparison of fits to KIC-8144222 photometry.

Cotrending	σ	90 %	99 %
2 PMs	24	28	53
4 PMs	24	28	53
Spline	53	66	146

Note. — Standard deviation, 90 percentile, and 99 percentile (in ppm) of the $\overline{\delta F}$ distributions for KIC-8144222 using different cotrending schemes. The spline fit is included as a baseline since it incorporates no ensemble-based cotrending information. In computing $\overline{\delta F}$, we have assumed a 12-hour transit duration. All cotrending approaches yield tighter $\overline{\delta F}$ distributions than the spline baseline.

We assess cotrending performance in the context of transit detectability. We note that PDC outputs are not directly used in transit detection. PDC light curves are subject to additional detrending (mostly of low frequency content) before the transiting planet search is run (Tenenbaum et al. 2010).

In Figure 6, we show fits to the KIC-8144222 photometry using 4 TERRA modes and the PDC algorithm. While PDC flattens photometry collected during the thermal transients, it injects high frequency noise into regions that are featureless in the TERRA-calibrated photometry. For KIC-8144222, the RMS scatter in the 12-hour $\overline{\delta F}$ distribution is 24 ppm for TERRA processed photometry and 36 ppm for PDC processed photometry.

Using 4 TERRA modes, we cotrend 100 stars selected at random from our 1000-star reference ensemble. We then compute 3, 6, and 12-hour $\overline{\delta F}$ from TERRA and PDC calibrated light curves. We then calculate the difference between the σ , 90%, and 99% statistics for TERRA and PDC cotrending. We show the distribution of these differences for the 12-hour $\overline{\delta F}$ in Figure 7. The median improvement in σ , 90%, and 99% using TERRA cotrending is 2.8, 6.6, and 8.7 ppm. We tabulate the median values of the σ , 90%, and 99% statistics in Table 2.

We believe that these comparisons are representative of the stars from which we constructed our reference ensemble ($12.5 < Kp < 13.5$ and $CDPP12 < 40$ ppm). These bright, low-noise stars are the most amenable to exoearth detection. Our comparisons do not pertain to stars with different brightness or noise level.

5. Conclusions

TERRA is a new technique for using ensemble photometry to self-calibrate instrumental systematics in *Kepler* light curves. We construct a simple noise model by running a high-pass filter and removing thermal settling events before computing principle components. For a typical $12.5 < Kp < 13.5$ and $CDPP12 < 40$ ppm star, TERRA produces ensemble-calibrated photometry with 33 ppm RMS scatter in 12 hour bins. With this noise level, a 100 ppm transit from an exoearth will be detected at $\sim 3\sigma$ per transit.

A potential drawback of removing thermal settling events is discarding photometry that contains a transit. Thermal settling events amounted to 14% of the valid cadences in Q1-Q8 photometry. Since signal to noise grows as the square root of the number of transits, removing 14% of the photometry results in a 7% reduction in the signal to noise of a given transit. The completeness of the survey may decrease slightly, since some borderline transits will remain below threshold. However, this can easily be overcome by gathering 14% more data.

Ensemble-based cotrending is most effective when the timescales in the ensemble are matched to the signal of interest. We are skeptical that a “one size fits all” approach exists and we encourage those who wish to get the most out of *Kepler* data to tune their cotrending to the timescale of

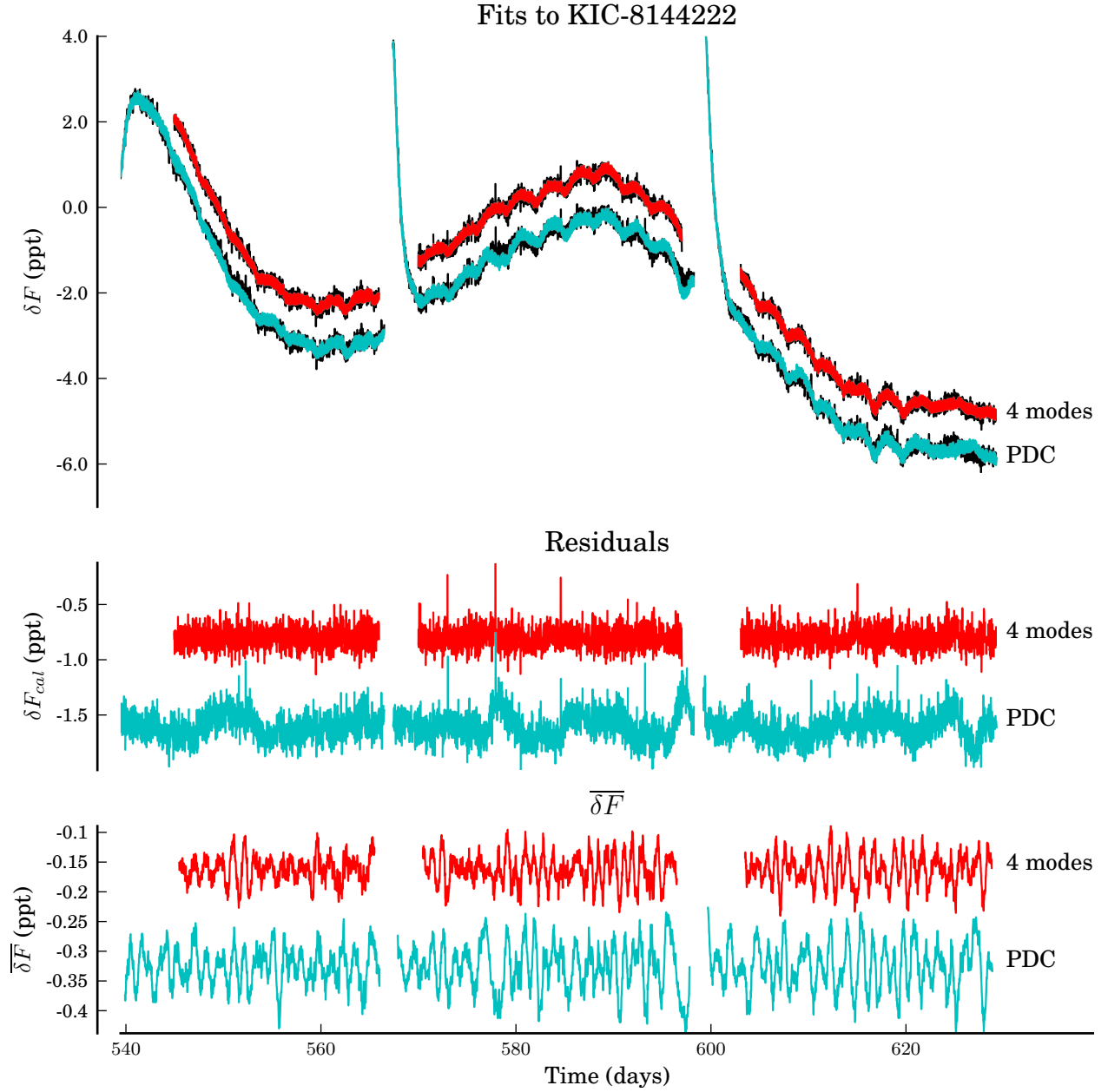


Fig. 6.— Same as Figure 5 except we compare fits using the 4 TERRA modes, with the PDC processed photometry. The bottom panel shows 12-hour $\overline{\delta F}$ where smaller scatter implies greater sensitivity to small transits. The RMS scatter in the 12-hour $\overline{\delta F}$ distribution is 24 ppm for TERRA processed photometry and 36 ppm for PDC processed photometry.

Table 2. Comparison of TERRA and PDC cotrending performance for 100 stars.

Transit Width (hours)	σ TERRA	σ PDC	90% TERRA	90% PDC	99% TERRA	99% PDC
3	58	60	68	76	129	141
6	43	45	50	57	97	105
12	33	37	39	47	76	88

Note. — A comparison of the $\overline{\delta F}$ distributions using TERRA and PDC cotrending of 100 stars drawn randomly from our 1000-star sample. We have assumed a range of transit widths. We show the median values of the standard deviation, 90 percentile, and 99 percentile (in ppm) of the $\overline{\delta F}$ distributions. For these 100 stars, TERRA yields tighter distributions of $\overline{\delta F}$. The improvement ranges from 8 to 12 ppm in the 99 % statistic.

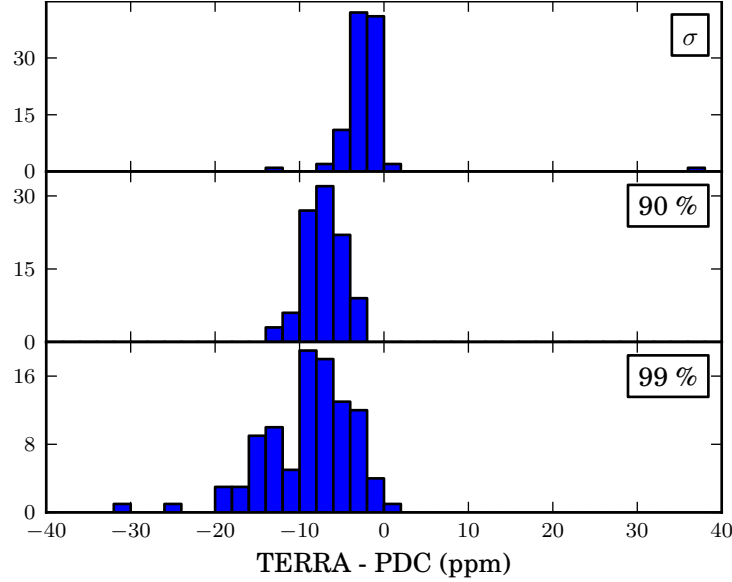


Fig. 7.— We computed the standard deviation, 90 percentile, and 99 percentile (in ppm) of 12-hour $\overline{\delta F}$ for 100 light curves using TERRA and PDC cotrending. The histograms show the difference of the TERRA and PDC statistics. Negative values mean a tighter $\overline{\delta F}$ distribution using our cotrending and hence a lower noise floor in a transit search.

their signals of interest.

The authors are indebted to Jon M. Jenkins, Andrew W. Howard, Douglas A. Caldwell, Thomas Barclay, Jeffrey C. Smith, and Jeffrey D. Scargle for productive and enlightening conversations that improved this work. We acknowledge salary support for Petigura by the National Science Foundation through the Graduate Research Fellowship Program. This work made use of NASA’s Astrophysics Data System Bibliographic Services as well as the SciPy (Jones et al. 2001–), IPython (Pérez & Granger 2007), and Matplotlib (Hunter 2007) Python modules.

REFERENCES

- Batalha, N. M., et al. 2011, *ApJ*, 729, 27
- . 2012, *arXiv*, 1202, 5852
- Borucki, W. J., et al. 2011, *ApJ*, 736, 19
- Christiansen, J. L., et al. 2011, *Kepler Data Characteristics Handbook* (KSCI-19040-002)
- Doyle, L. R., et al. 2011, *Science*, 333, 1602
- Fischer, D. A., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 419, 2900
- Fraquelli, D., & Thompson, S. E. 2012, *Kepler Archive Manual* (KDMC-10008-004)
- Fressin, F., et al. 2012, *Nature*, 482, 195
- Huang, X., Bakos, G. Á., & Hartman, J. D. 2012, *arXiv*, astro-ph.EP
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Jenkins, J. M., et al. 2010, *Software and Cyberinfrastructure for Astronomy*. Edited by Radziwill, 7740, 10
- Jones, E., Oliphant, T., Peterson, P., et al. 2001–, *SciPy: Open source scientific tools for Python*
- Kovács, G., Bakos, G., & Noyes, R. W. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 557
- Lintott, C., et al. 2012, *eprint arXiv:1202.6007*
- Lissauer, J. J., et al. 2011, *Nature*, 469, 53
- . 2012
- Morton, T. D., & Johnson, J. A. 2011, *ApJ*, 738, 170

- Muirhead, P. S., et al. 2012, *ApJ*, 747, 144
- Ofir, A., & Dreizler, S. 2012, arXiv, astro-ph.EP
- Pérez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, 9, 21
- Smith, J. C., et al. 2012, arXiv, astro-ph.IM
- Stumpe, M. C., et al. 2012, arXiv, astro-ph.IM
- Tamuz, O., Mazeh, T., & Zucker, S. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 1466
- Tenenbaum, P., Bryson, S. T., Chandrasekaran, H., Li, J., Quintana, E., Twicken, J. D., & Jenkins, J. M. 2010, *Software and Cyberinfrastructure for Astronomy*. Edited by Radziwill, 7740, 16
- Twicken, J. D., Chandrasekaran, H., Jenkins, J. M., Gunter, J. P., Girouard, F., & Klaus, T. C. 2010, *Software and Cyberinfrastructure for Astronomy*. Edited by Radziwill, 7740, 62
- Van Cleve, J. E., & Caldwell, D. A. 2009, *Kepler Instrument Handbook* (KSCI-19033-001)